

APPLICABILITY AND VALIDITY OF SOME
NUMERICAL METHODS IN PLANT ECOLOGY
I Principal components analysis

E. Kvillner

 Lund 1978

APPLICABILITY AND VALIDITY OF SOME NUMERICAL METHODS IN PLANT ECOLOGY

I Principal components analysis

E. Kvillner

APPLICABILITY AND VALIDITY OF SOME
NUMERICAL METHODS IN PLANT ECOLOGY
I Principal components analysis



E. Kullner

1 INTRODUCTION

In vegetation science a great number of different numerical methods have rapidly developed during the last decades. Most multivariate statistical methods belong to one of five classes: regression analysis, correlation analysis, components and factor analysis, test of homogeneity, classification and discriminant analysis (Dagnelie, 1971). Having a firm mathematical basis, the methods have been more or less successfully applied to biological data, which often are very complex in nature. However, the different procedures are often suitable for different purposes and usually involve different assumptions about the variables (see e.g. Cooley and Lohnes, 1962, Lawley and Maxwell, 1963, Seal, 1964, Austin and Orloci, 1966, Harman, 1967, Noy-Meir, 1973, Ihm and Groenewoud, 1975, Austin, 1976, Gauch and Wentworth, 1976, Gauch and Whittaker, 1976, Smartt, Meacock and Lambert, 1976, Feoli, 1977, Noy-Meir and Whittaker, 1977). Various coefficients analyzing the relationship among species and among stands have been suggested (see e.g. Noy-Meir, 1973, Noy-Meir, Walker and Williams, 1975).

But there are still uncertainties about the function of numerical methods. Different procedures, numerical and non-numerical, have both general and specific characteristics. If a method is to be useful then its advantages and limitations must be investigated. The simplest, but most important criterion, of all numerical methods is that they should give results which are independent of the user, i.e., they must be reproducible.

This paper deals with the applicability and validity of principal components analysis, PCA, and summarizes some of the main criticisms of the method. Two coefficients, covariance and correlation, are compared for their dependence on the relationship between presence of species and variance as well as for mean frequency and variance.

Since the PCA-ordination is based upon assumptions of linearity, the distortions of the ecological gradient, to which species respond, may be more or less pronounced by such a technique. This property of PCA, the so-called "horse-shoe effect", has been demonstrated on vegetation data by, e.g., Swan (1970). In many cases, the more "simple" the underlying gradient is assumed to be found, the more skewed it is actually found. This drawback has led to controversies over proper interpretations (see, e.g., Beals 1973). The PCA-procedure is however most useful when applied to a narrow range of the biological spectrum (Greig-Smith, 1964, Whittaker, 1967). It has therefore been suggested that a large-scale gradient may be divided into less distorted narrow ranges by use of some numerical classification method.

Usually the PCA-ordination is applied only to vegetation data. Some of the problems concerning the method are specific for vegetation data. When applied to environmental variables a less skewed structure is found. Hypotheses regarding the relationship between vegetation and environment may be generated by correlation between the PCA-structures of both systems (see, e.g., Andersson and Kvillner, 1977).

There is thus a poor compatibility of PCA with gradient analysis as, e.g., Nichols (1977) has provided. However, there is a variety of purposes for which PCA may be an useful technique, e.g., generation of hypotheses regarding physiological factors working on environmental gradients (see, e.g., Diaz, Garcia Nova and Merino, 1976) as well as reduction of the number of variables or samples. Some applications are discussed in this paper.

The PCA-ordination is here applied to two very different ecosystems, viz., the nutrient-deficient and elevated (ombrotrophic) mire complex of Stordalen, northern Sweden (Sonesson, 1970, Rosswall et al., 1975, Sonesson and Kvillner,

1977, Kvillner and Sonesson, 1977) and to a deciduous woodland and meadow area with a canopy of oak (*Quercus robur*), linden (*Tilia cordata*) and hazel (*Corylus avellana*) of Linnebjerg, southern Sweden (Andersson, 1970, Andersson and Kvillner, 1977). In this paper the materials are used only to demonstrate some major principles of PCA.

2 MATERIAL AND METHODS

2.1 Sampling

2.1.1 Vegetation

The sampling technique used at Stordalen was a point-frequency method (Greig-Smith, 1964, Sonesson, 1973). The vegetation data were sampled in 4 large quadrats, 50 x 50 m. In every quadrat 18 small quadrats, 2.5 x 2.5 m, were placed randomly. Ten frames, each consisting of 30 pins, 1.5 cm apart, were located at random in each of the small quadrats, thus making a total of 720 frames in 72 small quadrats (Johnsson, 1977). Presence/absence data for 36 species occurring in the field- and bottomlayer were thus collected. Two main computing units were used in the PCA-ordination, one based on the frequency of species collected in 30 pins in each frame (the "720"-material) and another based on the frequency of species calculated on 10 frames in each material (the "72"-material).

The sampling technique used at Linnebjerg was a contiguous quadrat method (Groenewoud, 1965, Andersson, 1970). The vegetation data were sampled in an area of 130 x 140 m, systematically divided into 14 transects, 130 m in length or 182 large quadrats, 10 x 10 m. The quadrats were systematically divided into 16 small quadrats, 2.5 x 2.5 m. Presence/absence data of 62 species occurring in the field-

layer were thus collected in every small quadrat. Two main computing units were used in the PCA-ordination, one based on the frequency of species collected in 16 small quadrats in each large quadrat (the "182"-material), another based on a subsample of this, where 97 large quadrats were selected (approximately every second transect) according to the environmental data sampling (the "97"-material).

2.1.2 Environment

Measurements were made on the humus layer of the Linnebjerg area of actual water content (weight %), pH_{KCl} , $\text{pH}_{\text{H}_2\text{O}}$, cation exchange capacity ($\text{meq} \cdot \text{dm}^{-3}$), base saturation (%), ^{25}C exchangeable hydrogen, calcium, potassium, magnesium, sodium and phosphorus ($\text{mmol} \cdot \text{dm}^{-3}$), total nitrogen ($\text{g} \cdot \text{dm}^{-3}$), loss on ignition (% of dry weight) and dry bulk density ($\text{g} \cdot \text{cm}^{-3}$), thus making a total of 14 environmental variables (for analytical methods, see Andersson, 1970). Of the original 182 large quadrats, 10 x 10 m, the centres of 97 quadrats in 8 transects out of the 14 transects available in the area were subjectively chosen for environmental data sampling.

2.2 Numerical methods

2.2.1 Covariance and correlation coefficients

Similarity between species is described in this paper by use of both covariance (cov) and correlation (r) coefficients. For any two species h and k the coefficients are calculated as

$$\text{cov}_{hk} = \frac{\sum_{i=1}^n [(X_{ih} - \bar{X}_h) (X_{ik} - \bar{X}_k)]}{n - 1} \quad \text{range indeterminate}$$

$$r_{hk} = \frac{\sum_{i=1}^n [(X_{ih} - \bar{X}_h) (X_{ik} - \bar{X}_k)]}{s_h \cdot s_k} \quad -1 \leq r \leq +1$$

where x_{ih} represents the quantitative value of species h in any stand i , $\bar{x}_h$ is the mean value of species h in all stands, x_{ik} is the quantitative value of species k in any stand i and $\bar{x}_k$ is the mean of species k in all stands. s_h and s_k represent the standard deviation of species h and k respectively.

Thus the covariance coefficient transforms the original species data by use of data centering $[(x_{ih} - \bar{x}_h)]$, but not

of data standardization $[\frac{(x_{ih} - \bar{x}_h)}{s_h}]$ as the correlation

coefficient does. Centering of species makes the contribution of each species proportional to its variance (Noy-Meir, 1973, Noy-Meir, Walker and Williams, 1975). With presence data, this means that species with 50 % presence in all stands will contribute the most. With quantitative data a similar pattern is discerned (c.f. however Noy-Meir, 1973). By use of the correlation coefficient, standardization by standard deviation equalizes the contribution of all species, abundant or rare. Thus the presence of a rare species is proportionally more important than that of an abundant one. The relationship between presence and variance (s_h^2 qual) and between mean frequency ($\bar{x}_h$) and variance (s_h^2 quant) is shown in Fig. 3 for the case of the Linnebjerg "72"-material.

2.2.2 Principal components analysis

There are three major steps in the PCA-ordination (Fig. 1), viz., (1) the preparation of the cross-products matrix, R (e.g., correlation and covariance matrix), i.e., calculating suitable coefficients between all pairs of species (R -technique) or between all pairs of stands (Q -technique), (2) the extraction of the initial factors/components, i.e., exploring possible data reduction during factorization and (3) the rotation to

a terminal solution, i.e., searching for simple and interpretable factors/components.

Although a correlation matrix is used in Fig. 1, the description is in most part applicable to the PCA-ordination of a covariance matrix too. All the diagonal elements of a correlation matrix are however equal to 1.00, while the diagonal elements of a covariance matrix are equal to the variances (s^2) of the variables, since variance can be thought of as the covariance of a variable with itself. Thus for the case of the R-technique, a species contributes proportionally to its variance in a covariance matrix, while the contributions of all species are equalized in a correlation matrix (c.f. p. 5).

During factorization, when the determinantal equation $|R - \lambda I| = 0$ is used, it is inherent that the rows of the matrix are linearly dependent, i.e., they are linear combinations of each other. The ordination model implicit is thus linear, i.e., the relationships among species and between species and ecological gradients are assumed to be linear. Two types of species response to a single gradient are outlined (Whittaker, 1967, Swan, 1970, Groenewoud, 1976), viz, S-shaped curves (i.e., species having their peak values at the end of the gradient) and bell-shaped curves (i.e., species having their peak values progressively closer to the centre of the gradient). As the gradient increases in range, the number of bell-shaped curves becomes larger. This means that only within restricted parts of the gradient can approximately linear relationships with species be assumed.

The effect of different overlapping species populations along the ecological gradient is a further important fact to be considered. Widely overlapping species decrease the ratio of bell-shaped to S-shaped curves (Swan, 1970). Swan further showed, using a PCA of an Euclidean distance matrix calcu-

lated between stands of simulated vegetation data, that even widely overlapping species populations give rise to a PCA-solution with serious distortions of the gradient, i.e., the original straight line of the ecological gradient bent into a complex, the so-called "horse-shoe effect", so that the most dissimilar species come to lie closely adjacent. With fewer overlapping species, the distortions become even greater. A decrease in population overlap is consequently accompanied by a progressively poorer representation of population changes along a gradient when examined by PCA-ordination.

A way to minimize the "horse-shoe effect" is to consider only a narrow range of the gradient. Thus an increase in population overlap will occur and consequently the ratio of bell-shaped to S-shaped curves will decrease, i.e., the relationship between species and the environmental gradient may approach the assumptions of linearity. The maximum variance extracted by a large-scale gradient, distorted by the PCA-ordination, still reveals the most important variation in the data but may also hide other variation obvious in the field along a narrow range of the gradient (Andersson and Kvillner, 1977). In fact, Greig-Smith (1964) suggests that factor analysis ordination (thus including PCA) may be most useful when applied to a narrow range of the ecological spectrum available for study.

Although the principal axes extracted are mathematically independent of each other (Gittins, 1968), i.e., they are orthogonal, they may not be ecologically independent since biological variables may not act independently. In fact, the interpretation of the first two principal axes of ordination of vegetation data is often easiest along the diagonals, i.e., the second principal axis may contribute more to the separation of loadings of variables or stand

scores than it does to a linear correlation with any specific environmental variable.

The main criticism of the use of PCA-ordination in vegetation analysis is that in most cases the assumption of linearity is ecologically incorrect. For environmental variables, the assumption may be more nearly correct. This will be discussed later on.

Some mathematical synonyms for terms used in matrix computation are given in Fig. 1. A more detailed explanation is given in, e.g., Cooley and Lohnes (1962), Lawley and Maxwell (1963), Groenewoud (1965) and Gittins (1968).

The R-technique is used for the PCA-ordinations of all materials in this paper. The vegetation data, i.e., the frequency of species, were used to calculate both covariance and correlation matrices, the environmental data to calculate correlation matrix only. The calculations were made on the Univac 1103 computer of the University of Lund. The standard program BMDP P4M was used.

3 RESULTS

Only a few of the many graphs prepared to test possible relationships are presented in this paper. The figures have been chosen to illustrate some major principles of the PCA-ordination. Ecological interpretations are presented in Kvillner and Sonesson (1977) and Andersson and Kvillner (1977).

The proportions of the maximum variance (λ) extracted by the principal axes of PCA-ordination of vegetation data differed both with regards to the coefficients used and to the various materials (Fig. 2). The axes of the PCA-ordina-

tions of covariance matrix of all the materials accounted for larger proportions of the total variance in comparison with the corresponding axes of PCA-ordinations of correlation matrix. By comparing different vegetation materials, the most prominent difference was found between the Stordalen "72"- and "720"-materials (curve D and E). While two axes were required to account for more than 30 % of the total variance in the "72"-material, ten axes were needed to reach the same limit in the "720"-material (correlation matrix, Fig. 2a). Only small differences were found by comparing the two vegetation materials of Linnebjerg (curve B and C). The first axis of the PCA-ordination of environmental data (curve A, Fig 2a) accounted for more than 60 % of the total variance.

In order to study the effect of the two coefficients, the Linnebjerg "97"-material is used as a general example. For qualitative vegetation data, the relationship between commonness of species (% presence of a species h in all stands) and variance ($s_{h \text{ qual}}^2$) of presence/absence data associated with each species shows a smooth pattern (Fig. 3a). The curve forms a complete arch with a peak at 50 % presence. For quantitative data, the relationship between mean frequency of species ($\bar{X}_h$) and variance ($s_{h \text{ quant}}^2$) is not so clear though still obvious (Fig. 3b). However, neither the upper limit of the curve nor the magnitude of decrease could be stated, because of the few observations with mean frequency > 40 %. Other materials do show the same general shape as that shown in Fig. 3b. The peak is at a mean frequency of 50 %, but the amplitude of the variance differs with each material.

These relationships are factors limiting the magnitude of covariance between two species. High covariance can accordingly occur between two species only when both are relatively common and when they display the same phytosociological pattern (c.f. Goff, 1975).

There is a correspondence between commonness and ordination position of species (i.e., the loadings of species on the principal axes, Fig. 4). In the PCA-ordination of the covariance matrix, the most abundant species were located near the periphery, thus playing an important role in the location of the axes, while the rare species were clustered near the origin (Fig. 4a). In the PCA-ordination of the correlation matrix, the rare species were given a relatively greater significance (Fig. 4b). Species with a frequency up to 20 % had, by comparison with the covariance model, considerable influence on the development of the axes.

4 DISCUSSION

Ordination - mathematical elegance and ecological naivité ?
(c.f. Beals, 1973)

The primary task of ordination is to express a large number of observed variables (species or environmental variables) in terms of a smaller number of hypothetical, non-observed and new variables (principal axes or components) thereby expressing the maximum variance in the original data. Whether these new axes have any meaning apart from the mathematical assumptions of the model of construction depends on the nature of the original variables. It is usually assumed that merely having extracted the maximum variance, this "ought" to have some ecological meaning in relation to the variables being studied (Austin, 1968).

The ordination technique is designed to extract the principal axes in order of decreasing importance (i.e., variance, λ), and the lengths of the first say two or three axes are consequently often used as a measure of the "efficiency" of the ordination. This "efficiency" is however dependent upon the coefficients being used (Fig. 2). The PCA-ordinations

of the different materials used are all similar with respect to decreasing variance with increasing numbers of axes. As stated by Sundman and Gyllenborg (1967), the more homogeneous the biological material to be described by an ordination, the steeper is the curve of cumulative variance. If this hold true, the environmental material is more homogeneous than any of the vegetation materials (Fig. 2a). The Linnebjerg "97"- and "182"-materials as well as the Stordalen "72"-materials are consequently more "homogeneous" than the Stordalen "720"-material. Using the criterion of efficiency, the use of a covariance matrix seems usually to produce the most "efficient" PCA-ordinations of the vegetation materials.

The explanation of these differences between the environment and the vegetation materials on one hand and between the different vegetation materials on the other must be attributable partly to the nature of the variables themselves and partly to the ability of the ordination model, based as it is on assumptions of linearity, to describe the nature of the different variables.

The weighting of species, i.e., their qualitative and quantitative values, has an important effect during computation, but the significance of sampling technique may often be underestimated. Geometrically, the principal axes are the major axes of the point cluster of species or stands, i.e., their origin lies at the centre of gravity (Fig. 1), and they are therefore dependent of the shape of the cluster, which in turn depends on the choice, availability and sampling of the stands. If one part of the cluster is "oversampled", the axes will reflect this. The results of different PCA-ordinations are consequently difficult to compare if different sampling techniques have been used, because a "satisfying" result or a high "efficiency" depends upon many causes, some of which have already been indicated.

Assumptions have been made that species variance ($s_{h \text{ quant}}^2$) tends to increase with the mean ($\bar{x}_{h \text{ quant}}$) and that species with a high variance usually also will be the most abundant ones (Noy-Meir, 1973, and Noy-Meir, Walker and Williams, 1975). There are however some important limitation of this increase (Fig. 3). For qualitative as well as for quantitative species data the relationship seems to be curvilinear with a peak at 50 % presence/frequency. This must be kept in mind when decisions on proper coefficients are made. The use of the correlation coefficient removes the dependence of the variance on the mean (c.f. p. 5). Instead rare species will have a significant influence on the development of axes (Fig. 4). Species positioned near the origin in a PCA based on a covariance matrix may have an important ecological meaning but their variation is overwhelmed by others. These situations may not always be desirable and considerations must be laid upon appropriate studies on infrequent as well as frequent species. One way to make a choice of coefficients to measure species relationship, would be to consider the number of common and rare species. If there are more common species than rare ones, the covariance coefficient may be more valuable, in the opposite case the correlation coefficient may be favourable. In Noy-Meir (1973) and Noy-Meir, Walker and Williams (1975) some other coefficients are considered. However, for all data transformations, special care must be taken not to overwhelm the PCA-analysis by purely numerical effects (c.f. Nichols, 1977).

The degree of presence of a species can always be recorded, but there is no corresponding measure of the degree of absence (see however Swan, 1970). A large number of zero values, i.e., absence of a species, will strongly affect the PCA-ordination (like any other numerical method) while environmental variables can be recorded along the whole of the ecological gradient. They are usually more or less normally distributed or at least they can be transformed to, e.g., a logarithmic scale if the distribution on linear scale is skewed. These facts may be the

most abundant ones when considering the maximum variance extracted in PCA-ordinations of environmental data. Thus the use of PCA-ordination on environmental data is more likely to be "satisfactory".

Gradual changes in the plant cover can be described as the variation along vegetational gradients (Sjörs, 1950), which in turn relate to changes in the environment. Thus it seems likely that the relations among species can be summarized in terms of a few underlying variables as the mathematical construction of principal axes implies. Any environmental variable can be later compared with these patterns to search for correlated features. The expectations of receiving an useful tool to reconcile and study ecological gradients have often been further strengthened by the association of PCA as "the gradient analysis method" (c.f. Whittaker, 1967: indirect gradient analysis). Failure to reconstruct assumed structure with actual PCA-structure owing to the "horse-shoe effect" has therefore led to the recent controversies over proper interpretations (c.f. Gauch and Whittaker, 1972, Beals, 1973).

The low variance accounted for by the first axes of the PCA-ordination of the Stordalen "720"-material (curve E in Fig. 2) may thus be attributable to the "horse-shoe effect", which strongly distorts the underlying gradients. The interpretation of such a PCA-ordination is therefore limited. One of the primary vegetational gradients in the mire vegetation, viz., the "hummock--mud-bottom" gradient, can be found but in a rather complex way. This gradient is fairly easily followed in the Stordalen "72"-material, where the large-scale vegetational gradient "mire margin - mire expanse" is obscure (Kvillner and Sonesson, 1977). The greatest difference between the two vegetation materials of Stordalen depends on the different sampling sizes used, which for the "720"-material may be greater than the appropriate "minimum" size.

The major vegetational gradient "low herb - tall herb", dividing the species into woodland and meadow species, is found along the diagonals of the first two axes of PCA-ordination of the Linnebjerg "97"- and "182"-materials (Andersson and Kvillner 1977). The ecological gradient "dry - wet", non-numerically recognized from considerations of the variation in soil moisture and soil aeration was skewed. An ecological gradient, numerically recognized in the environmental data from considerations of acid-base status, soil moisture and stores of exchangeable base cations, is comparatively less distorted. The vegetational gradient "tall herb - sedge" non-numerically recognized in the open meadow vegetation, is but partly recognized on the principal axes of the PCA-ordination. However, some of the distortions of the gradient may be due to the large sample size. Some of the important variation is thereby hidden and the degree of resolution is decreased.

The ability of PCA-ordination is thus limited. Although skewed, the PCA-structure may still be an effective tool in studying community structure, provided the aim is to reduce and simplify complex vegetation data. Since crucial experiments are difficult to conceive, as is usually the case when studying relationships among species and environment on a community level, a broader definition of ordination is needed as "any arrangement summarizing relationships among items" (Nichols, 1977).

Despite the effect of assumptions of linearity, the PCA-ordination is still an exploratory technique of the greatest value especially in the initial stages of a study of vegetation. The main assumption is that a latent structure or organisation of some kind exists within the observation set. If it were not so then any vegetation study would be aimless. In view of this, ordination should be considered as an useful means of simplifying the description of the vegetation in

such a way, that the generation of hypotheses regarding the causal environmental factors is stimulated and simplified. The great value of simultaneous ordination of environmental and vegetational data should be emphasized.

There are however some further capabilities of PCA not fully considered in plant ecology until recently. The reducing power of PCA-ordination can be used, e.g., to extract a representative number of samples or variables for more detailed investigations (Nilsson, 1977). Summarization of large species correlation matrices can be used to identify "ecological groups" (Noy-Meir, 1973a). Graphical presentations of species loadings on the principal axes have been developed, e.g., by Goff (1975). PCA-ordination may thus be an useful technique for a wide variety of purposes.

ACKNOWLEDGEMENTS

For guidance and assistance I am deeply indebted to Professor N. Malmer, Department of Plant Ecology, University of Lund, and Dr. R.S. Clymo, Westfield college, University of London. Comments on the manuscript were given by Professor M. Sonesson, Drs. M. Danielsson, I. Nilsson, G. Regnell, I. Stjernquist and B. Wiman, Department of Plant Ecology, University of Lund, Dr F. Andersson, Swedish Coniferous Forest Project, and Dr. T. Rosswall, Institute of Microbiology, the Swedish University of Agriculture Sciences, Uppsala. Guidance during data processing was given by L. Robertsson, Lunds Datacentral, Lund. All graphs were drawn by Mrs. M. Varga. I wish to thank them all for valuable help and advice.

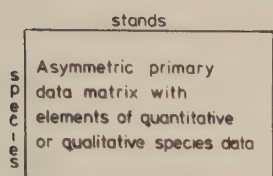
REFERENCES

- Andersson, F. 1970. Ecological studies in a Scanian woodland and meadow area, southern Sweden. I. Vegetational and environmental structure. - *Opera Botanica* 27. 190 pp.
- Andersson, F. and Kvillner, E. 1977. Ecological studies in a Scanian woodland and meadow area, southern Sweden. IV. Gradient analysis of vegetational and environmental patterns. - In press.
- Austin, M.P. 1968. An ordination study of a chalk grassland community. - *J.Ecol.* 56:739-757.
- Austin, M.P. 1976. On non-linear species response models in ordination. - *Vegetatio*, Vol. 33, 1:33-41.
- Austin, M.P. and Orloci, L. 1966. Geometric models in ecology. II. An evaluation of some ordination techniques. - *J. Ecol.* 54:217-227.
- Beals, E. 1973. Ordination: Mathematical elegance and ecological naivete? - *J. Ecol.* 61:23-35.
- Cooley, W.W. and Lohnes, P.R. 1962. *Multivariate Data Analysis*. - New York.
- Dagnelie, P. 1971. Some ideas on the use of multivariate statistical methods in ecology. - In: *Statistical ecology. Volume 3: Populations, Ecosystems and System Analysis*. p. 167-180. Patil, Pielou and Waters (eds.).
- Diaz, L.R., Garcia Nova, F. and Merino, J. 1976. On the ecological interpretation of principal components in factor analysis. - *Oecol. Plant.* 11:137-141.
- Feoli, E. 1977. On the resolving power of principal components analysis in plant community ordination. - *Vegetatio*, Vol. 33, 2/3:119-125.
- Gauch, H.G. and Whittaker, H. 1972. Comparison of ordination techniques. - *Ecol.* 53:868-875.
- Gauch, H.G. and Wentworth, T.R. 1976. Canonical correlation analysis as an ordination technique. - *Vegetatio*, Vol. 33, 1:17-22.
- Gauch, H.G. and Whittaker, 1976. Simulation of community patterns. - *Vegetatio*, Vol. 33, 1:13-16.
- Gittins, R. 1968. The Application of Ordination Techniques. - In: Rorison (ed), *Ecological Aspects of the Mineral Nutrition of Plants*. Symp. Brit. Ecol. Soc. 1968, 9:37-66.

- Goff, F.G. 1975. Comparison of species ordinations resulting from alternative indices of interspecific association and different numbers of included species. - *Vegetatio*, Vol. 31, 1:1-12.
- Greig-Smith, P. 1964. Quantitative plant ecology. - London. 246 pp.
- Groenewoud, H. van 1965. Ordination and classification of Swiss and Canadian coniferous forests by various biometric and other methods. - *Ber. Geobot. Forsch. Inst. Rübel* 36:28-102.
- Groenewoud, H. van 1976. Theoretical considerations on the covariation of species along ecological gradients with regards to multivariate analysis. *J. Ecol.* 64:837-847.
- Harman, H.H. 1967. Modern factor analysis. 2nd edn. - Chicago.
- Ihm, P. and Groenewoud, H. van 1975. A multivariate ordering of vegetation data based on Gaussian type gradient response curves. - *J. Ecol.* 63:767-777.
- Johnsson, S. 1977. Introduction. In: Sonesson, M. (ed.) 1977. Ecology of a Subarctic Mire. *Ecological Bulletins* Vol. 27.
- Kvillner, E. and Sonesson, M. 1977. Ordination of plant distribution and environmental data. - In: Sonesson, M. (ed.) 1977. Ecology of a Subarctic Mire. *Ecological Bulletins* Vol. 27.
- Lawley, D.N. and Maxwell, A.E. 1963. Factor analysis as a statistical method. - London.
- Nichols, S. 1977. On the interpretation of principal components analysis in ecological contexts. - *Vegetatio*, Vol. 34, 3:191-197.
- Nilsson, I. 1977. The acidification sensitivity of some Swedish forest soils. - *Meddn. Avd. Ekol. Bot., Lunds Univ.* 32. Lund.
- Noy-Meir, I. 1973a. Data transformation in ecological ordination. I. Some advantages of non-centering. - *J. Ecol.* 61:329-341.
- Noy-Meir, I. 1973b. Divisive polythetic classification of vegetation by optimized division on ordination components. - *J. Ecol.* 61:753-760.
- Noy-Meir, I., Walker, D. and Williams, W.T. 1975. Data transformation in ecological ordination. II. On the meaning of data standardization. *J. Ecol.* 63:729-800.
- Noy-Meir, I. and Whittaker, R.H. 1977. Continuous multivariate methods in community analysis: some problems and developments. - *Vegetatio*, Vol. 33, 2/3:79-98.

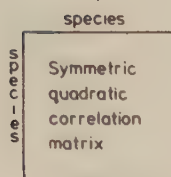
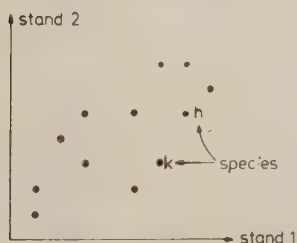
- Rosswall, T., Flower-Ellis, J.G.K., Johansson, L.-G., Jonsson, S., Rydeñ, B.E. and Sonesson, M. 1975. Word model of the Stordalen site. In: Heal, O.W. and Rosswall, T. (eds.).
- Seal, H. 1964. Multivariate statistical analysis for biologists. - London.
- Sjörs, H. 1950. Regional studies in North Swedish mire vegetation. - Botaniska notiser 1950:173-222.
- Smartt, P.F.M., Meacock, S.E. and Lambert, J.M. 1976. Investigations into the properties of quantitative vegetational data. II. Further data type comparisons. - J. Ecol. 64:41-78.
- Sonesson, M. 1970. Studies on the Mire vegetation in the Torne-träsk Area, Northern Sweden. III. Communities of the poor mires. - Opera Botanica 26. 120 pp.
- Sonesson, M. 1973. Point frequency determinations, Stordalen 1972. - In: Sonesson, M. (ed.) Progress Report 1972. IBP Swedish Tundra Project, Tech. Rep. 14.
- Sonesson, M. and Kvillner, E. 1977. Plant communities - a comparison between numerical and non-numerical classification methods. - In: Sonesson, M. (ed.) 1977. Ecology of a Subarctic Mire. Ecological Bulletins. Vol. 27.
- Sundman, V. and Gyllenberg, H.G. 1967. Application of factor analysis in microbiology. I. General aspects on the use of factor analysis in microbiology. - Ann. Acad. Sci. Fenn. A IV 112, 1-32.
- Swan, J.M.A. 1970. An examination of some ordination problems by use of simulated vegetation data. - Ecology 51:89-102.
- Whittaker, R.H. 1967. Gradient analysis of vegetation. - Biol. Rev. 49:208-264.

COMPUTATIONAL PROCEDURE

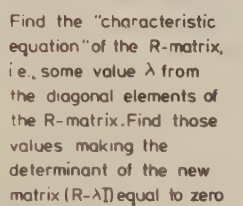
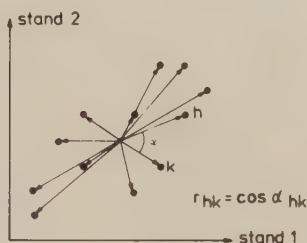


PRIMARY FIELD DATA

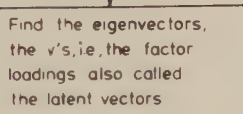
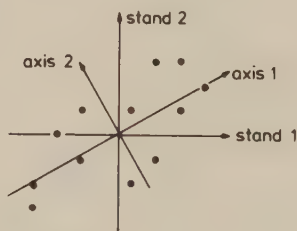
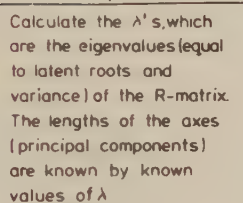
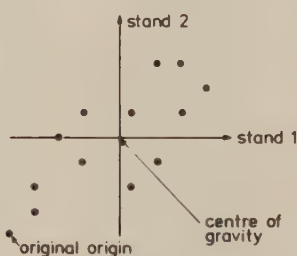
GEOMETRICAL VERSION



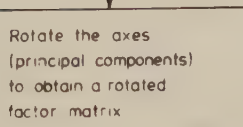
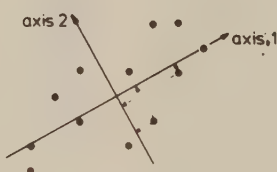
CORRELATION MATRIX R



FACTORIZATION
 $|R - \lambda I| = 0$



$(R - \lambda I) \cdot V = 0$



ROTATION

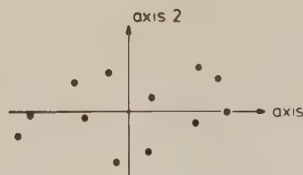


Fig. 1. Major steps in the R-technique of principal components analysis applied to quantitative data of twelve species on two stands. The geometrical version is illustrated together with some mathematical synonyms for matrix computation.

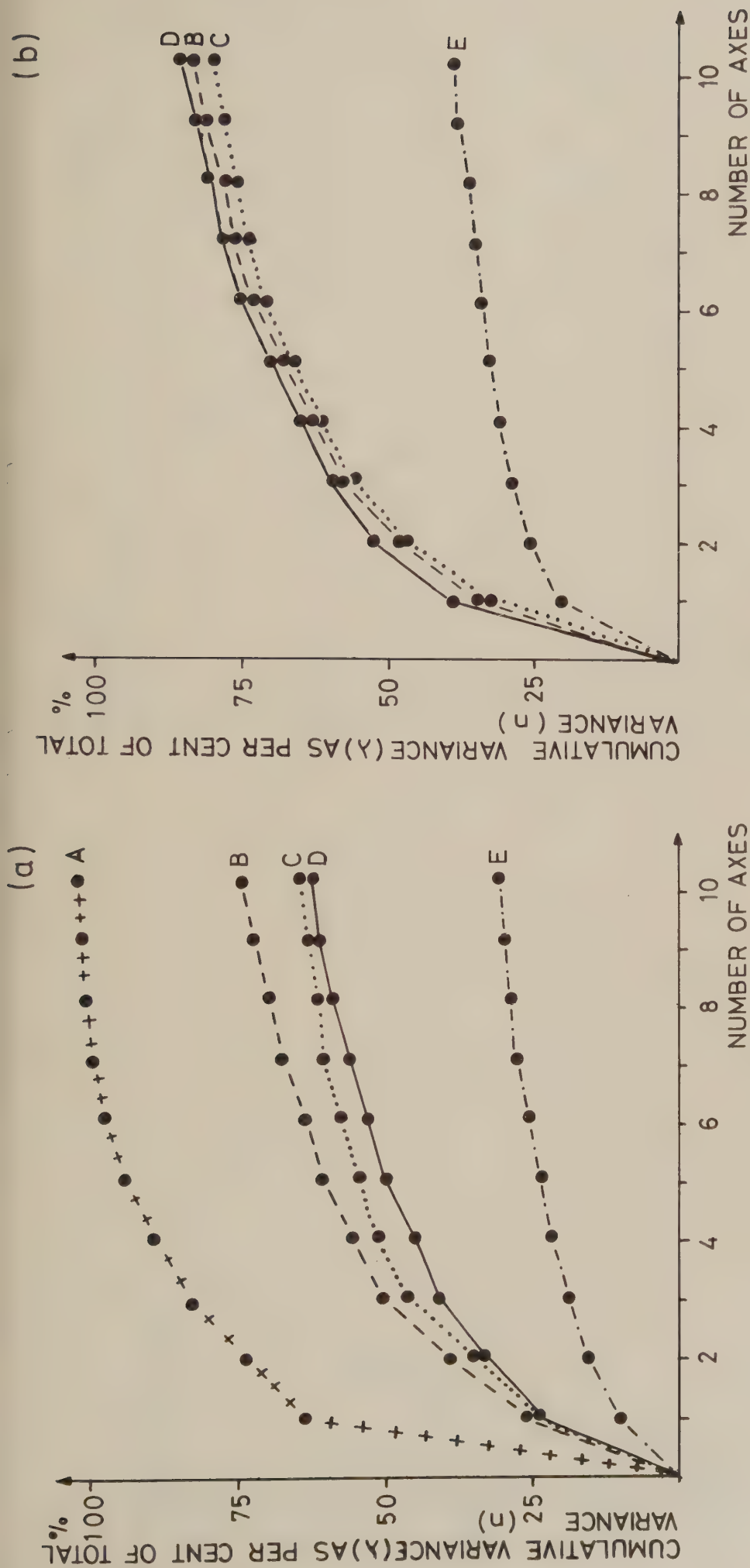


Fig. 2. Cumulative contribution of the variance (λ) to the total variance in the data as a function of the number of axes. (a) PCA-ordination of the covariance matrix, (b) PCA-ordination of the correlation matrix.

- A: $N = 97$, $n = 14$. PCA-ordination of environmental data of Linnebjör
 B: $N = 97$, $n = 62$. PCA-ordination of vegetation data of Linnebjör, the "97"-material
 C: $N = 182$, $n = 62$. PCA-ordination of vegetation data of Linnebjör, the "182"-material
 D: $N = 72$, $n = 36$. PCA-ordination of vegetation data of Stordalen, the "72"-material
 E: $N = 720$, $n = 36$. PCA-ordination of vegetation data of Stordalen, the "720"-material

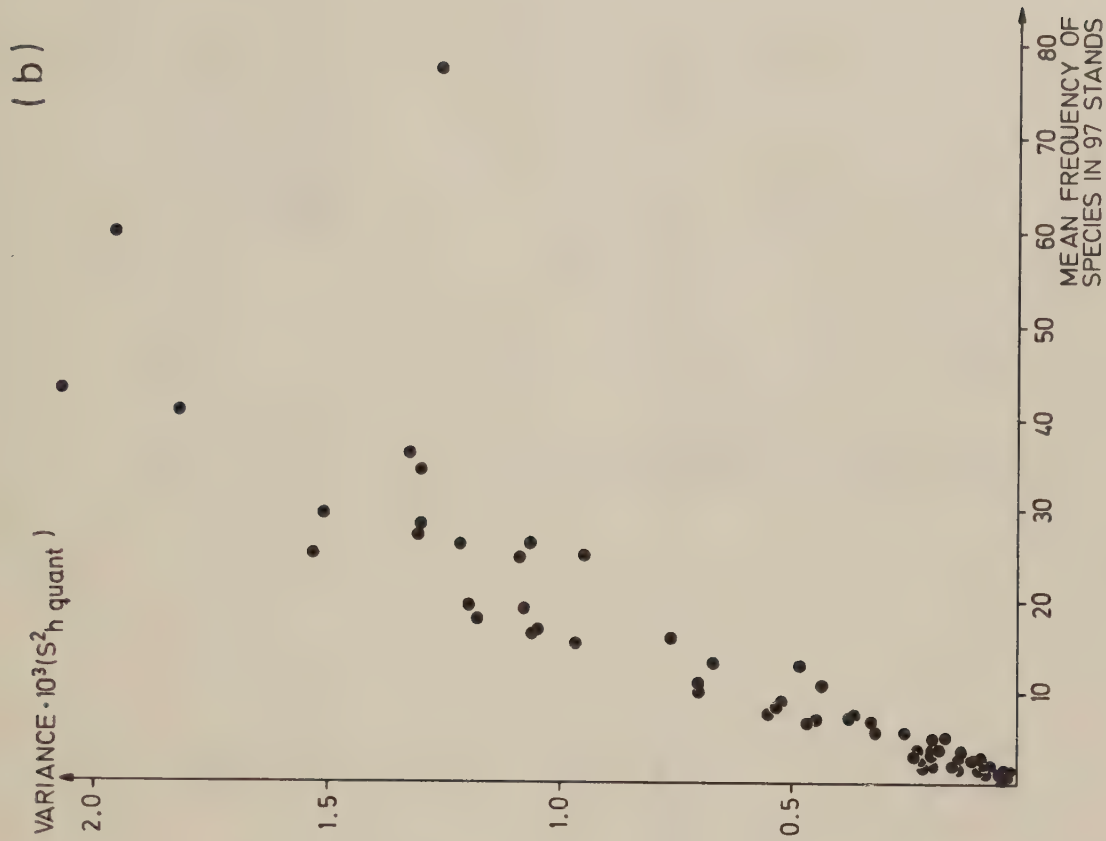
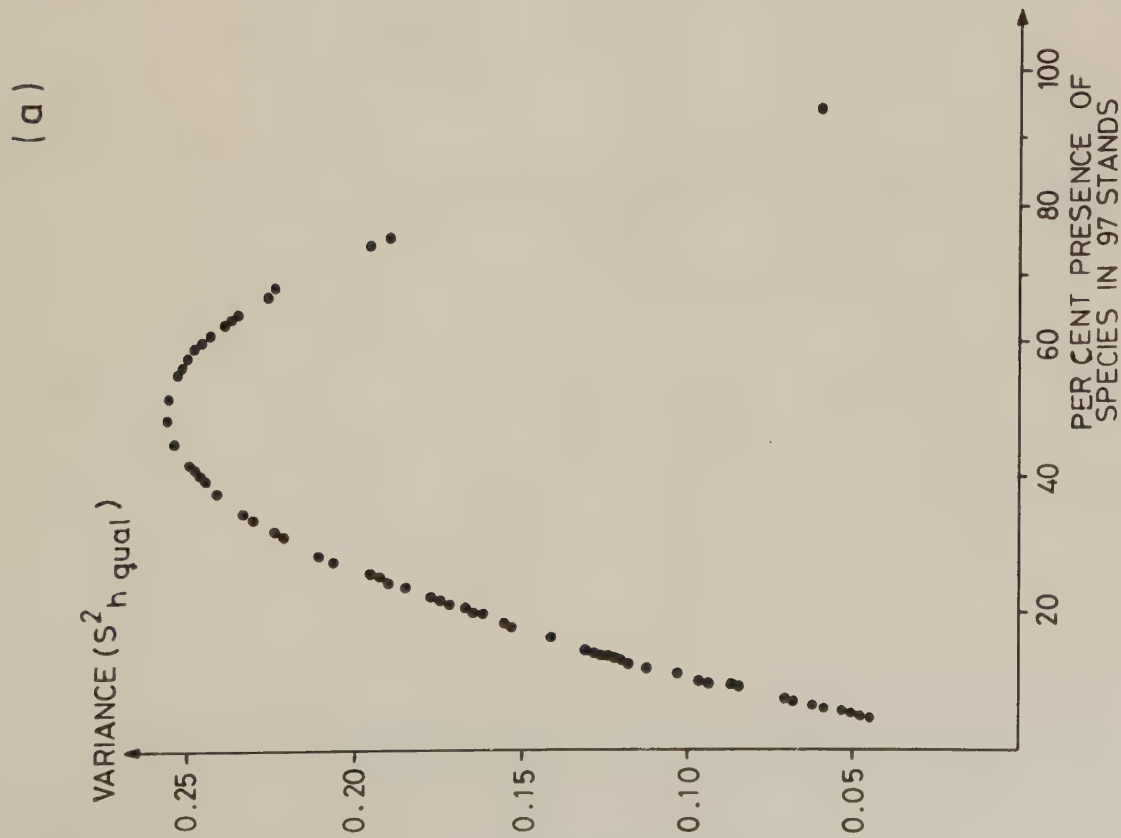


Fig. 3a. Relationship between presence of species and variance associated with the presence/absence data of the Linnebjør "97"-material.

Fig. 3b. Relationship between mean frequency of species ($\bar{x}_h$) and variance ($s^2_{h \text{ quant}}$) of the Linnebjør "97"-material.

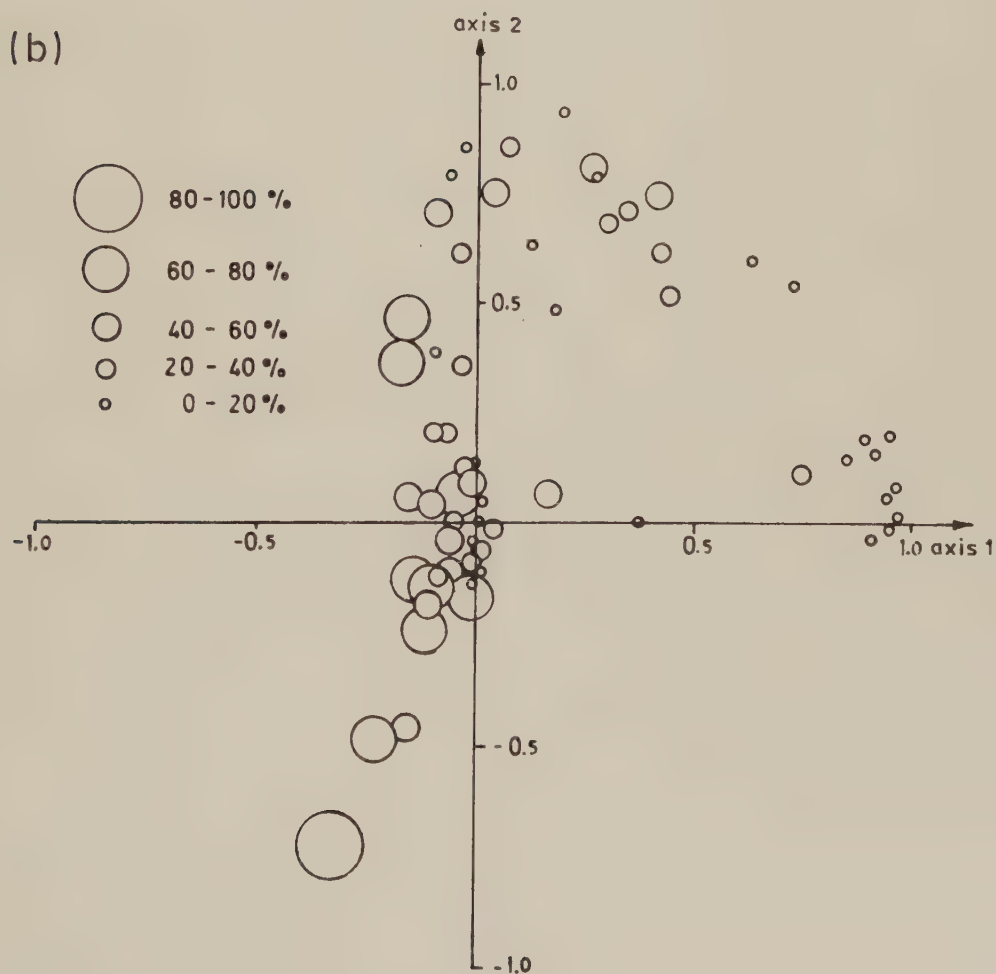
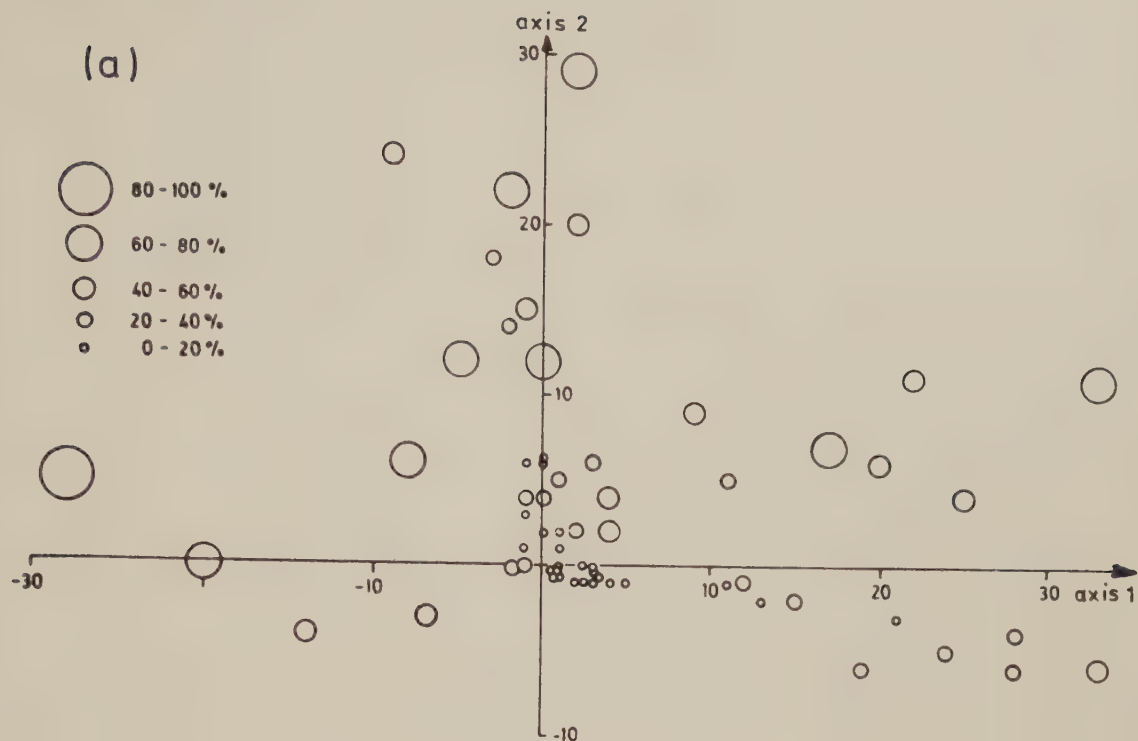


Fig. 4. Loadings of species on the first and second principal axes of PCA-ordination of correlation (a) and covariance matrix (b) of the Linnebjerg 97-material. % presence of a species in all 97 stands is indicated by the size of the circles.

